# yang_seonhyeHW9

*Seonhye Yang*

*2/8/2019*

```r
library(data.table)
hybrid_reg <- fread("http://users.stat.ufl.edu/~winner/data/hybrid_reg.csv")
hybrid <- data.table("hotel_energy.csv", header = TRUE)
```

## Question 1

```r
attach(hybrid_reg, warn.conflicts = F)
model <- lm(msrp~mpgmpge+accelrate)
model1 = confint(model,level=.95)
model1
```

```
##                   2.5 %     97.5 %
## (Intercept) -26628.4764 2008.72023
## mpgmpge       -277.3895   14.43946
## accelrate     3828.8331 5651.44832
```

```r
summary(model)
```

```
##
## Call:
## lm(formula = msrp ~ mpgmpge + accelrate)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -38435  -8709  -2836   7755  51093
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -12309.88    7246.60  -1.699   0.0914 .
## mpgmpge       -131.48      73.85  -1.780   0.0770 .
## accelrate     4740.14     461.21  10.278   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15330 on 150 degrees of freedom
## Multiple R-squared:  0.4945, Adjusted R-squared:  0.4878
## F-statistic: 73.37 on 2 and 150 DF,  p-value: < 2.2e-16
```

We are 95% confident that $\beta_{mpgmpge}$ falls between -277.38 and 14.43946. Additionally,We are 95% confident that $\beta_{areaccelrate}$ falls between -3828.8331 5and 651.44832. This means that 95% of all cars are affected in relation to MPGe and Acceleration by this range of values

-12309.88 + 4740.14*4500 ## Question 2

```r
attach(hybrid_reg, warn.conflicts = F)
xstar=data.frame(mpgmpge = 40, accelrate = 10)
predict(model,xstar,interval = "confidence",level = .99)
```

```
##         fit      lwr      upr
## 1 29832.53 25890.53 33774.53
```

```r
x1star=data.frame(1, mpgmpge = 40, accelrate = 10)
predict(model,x1star,interval = "prediction",level=.99)
```

```
##         fit       lwr     upr
## 1 29832.53 -10360.25 70025.3
```

Looking at the first interval, we can see that the lower and upper bounds are 25890.53 and 33774.53 respectively, meaning that for a car with 40 MPGe and an acceleration of 10, the price falls within that range. Looking at our predicted ranges, we can see that the lower bounds are -10360.25 and 70025.3, indicating that our predictor fits well for creating a range of prices based on these values.

## Question 3

```r
modelcoef<-coef(summary(model))
modelcoef
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) -12309.878  7246.6014 -1.698711 9.144694e-02
## mpgmpge       -131.475    73.8469 -1.780373 7.703926e-02
## accelrate     4740.141   461.2101 10.277616 4.336620e-19
```

```r
tval1<-(modelcoef[3, 1]-4500)/modelcoef[3, 2];tval1
```

```
## [1] 0.5206753
```

```r
dffit2<-model$df.residual;dffit2
```

```
## [1] 150
```

```r
pval=pt(tval1,dffit2,lower=F)
pval
```

```
## [1] 0.3016804
```

We obtain a p value of 0.3016, which is greater than 0.05, so we cannot reject the null hypothesis and must accept our hypothesis to be significant

## Question 4

```r
model3=lm(msrp~1, data = hybrid_reg, offset=-120*mpgmpge+4500*accelrate)
anova(model3,model)
```

```
## Analysis of Variance Table
##
## Model 1: msrp ~ 1
## Model 2: msrp ~ mpgmpge + accelrate
##   Res.Df        RSS Df Sum of Sq      F Pr(>F)
## 1    152 3.5357e+10
## 2    150 3.5257e+10  2 100566428 0.2139 0.8076
```

Looking at the ANOVA results, we can see that there is a computed P value of 0.2139, which is greater than 0.05. This result implies that we cannot reject the null hypothesis of H1:H0 is false.

## Question 5

We can start with $var(y_* - \hat{y}_*)$:

$$var(y_*) - var(\hat{y}_*)$$

$$var(\mathbf{X}_*^T \beta + \epsilon_*) - var(\hat{y}_*)$$

We know that $var(\mathbf{X}_*^T \beta + \epsilon_*) = \sigma^2$, so:

$$\sigma^2 - var(\hat{y}_*)$$

$$\sigma^2 - (\mathbf{X}_*^T)^2 var(\hat{\beta})$$

$$\sigma^2 - (\mathbf{X}_*^T)^2 (\sigma^2 (\mathbf{X}^T \mathbf{X})^{-1})$$

Rearranging this, we get:

$$\sigma^2 - (\mathbf{X}_*^T)(\sigma^2 (\mathbf{X}^T \mathbf{X})^{-1})(\mathbf{X}_*)$$

Factoring:

$$\sigma^2 (1 + \mathbf{X}_*^T (\mathbf{X}^T \mathbf{X})^{-1}) \mathbf{X}_*)$$

Which is equivalent to the solution we were looking for.